④

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS

AD-A203 081

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release; distribution unlimited

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION Research Laboratory of Electronics Massachusetts Institute of Technology	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) 77 Massachusetts Avenue Cambridge, MA 02139		7b. ADDRESS (City, State, and ZIP Code)

8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-82-K-0727

8c. ADDRESS (City, State, and ZIP Code) Department of the Navy Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO.	PROJECT NO. NR049-542	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Speech Recognition: Acoustic-Phonetic Knowledge Acquisition and Representation.

12. PERSONAL AUTHOR(S) Dr. Victor W. Zue

13a. TYPE OF REPORT End of Fiscal Year	13b. TIME COVERED FROM 10/1/88 TO 9/30/88	14. DATE OF REPORT (Year, Month, Day) 9/30/88	15. PAGE COUNT 9 pp.

16. SUPPLEMENTARY NOTATION

.

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

S DTIC ELECTED
NOV 1 0 1988
D ᶜᵛ D

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Barbara Passero RLE Contract Reports	22b. TELEPHONE (Include Area Code) (617) 253-2566	22c. OFFICE SYMBOL

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

"END OF FISCAL YEAR" REPORT

Speech Recognition:
Acoustic- Phonetic Knowledge
Acquisition and Representation

Office of Naval Research
Contract N00014-82-K-0727

Covering the Period
1 October 1987- 30 September 1988

Submitted by:

Victor W. Zue

Accesion For		
NTIS CPA&I		☑
DTIC TAB		☐
Unannounced		☐
Justific:		
By		
Distrib		
A		
Dist		

ONR Scientific Officer:

Dr. Alan Meyrowitz

30 September 1988   A-1

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Research Laboratory of Electronics
Cambridge, Massachusetts 02139

88 11 09 066

A. Description of Scientific Research Goals

Our long-term research goal is to develop and implement speaker-independent continuous speech recognition systems. We believe that the proper utilization of speech-specific knowledge is essential for such advanced systems. Our research is thus directed toward the acquisition, quantification, and representation of acoustic-phonetic and lexical knowledge, and the application of this knowledge to speech recognition algorithms. In addition, we are exploring new speech recognition alternatives based on AI and connectionist techniques.

B. Significant Results in the Last Year:

- We developed a statistical model for predicting the acoustic realization of stop consonants in various positions in the syllable template. A unification-based grammatical formalism was developed for incorporating this model into the lexical access algorithm. We provided an information-theoretic justification for the hierarchical structure of the syllable template.

- We analyzed segmented duration for vowels and fricatives in continuous speech. Based on contextual information, we developed durational models for vowels and fricatives that account for over 70% of the variance, using data from multiple, unknown speakers.

- We rigorously evaluated the ability of human spectrogram readers to identify stop consonants spoken by many talkers and in a variety of phonetic contexts. Incorporating the declarative knowledge used by the readers, we developed a knowledge-based system for stop identification. We achieved comparable system performance to that of the readers.

- We developed a technique for phonetic classification using artificial neural nets (ANN). Vowel classification accuracy was achieved, ranging from 66 to 100% under varying experimental conditions.

C Plans for Next Year's Research:

- We will complete the syllable-based lexical access model, and evaluate its effectiveness.

- We will develop durational models for other classes of speech sounds. We will investigate the effect of speaking rate on these models. We will use these results to develop a comprehensive model for segmental duration to aid speech recognition.

- We will refine our ANN-based classification procedures for phonetic classification, and evaluate their performance against that of more tradition techniques.

- We will investigate the role played by prosody on phonetic recognition and lexical access.

- We will investigate various alternatives for adaptation to improve recognition system performance.

D. Participants:

Principal Investigator
Victor W. Zue

Research Staff
David Kaufman
Michael Phillips
Stephanie Seneff

Graduate Students
Susan R. Dubois
Lori Lamel   (PhD degree granted May 1988)
Andrew Howitt   (S.M. degree granted August 1987)
Hong C. Leung
John F. Pitrelli
Mark Randolph

Undergraduate Students
Charles Jankowski   (S.B. degree granted May 1988)
Hirak Mitra   (S.B. degree granted May 1988)
David Whitney
Davin C. Wong

E Other sponsored research

Title: Acoustic-Phonetics Based Speech Recognition
Sponsor: Naval Electronic Systems Command (DARPA)
Amount: $2,018,533.00
Contract Period: 8 February 1985- 31 January 1989

F. Electronic mail address: Zue@goldilocks.mit.edu
                                         or
                                         Zue@mc.lcs.mit.edu

Appendix to follow

2

1. Papers Published in Refereed Journals:

Glass, J.R., and V.W. Zue, "Multi-Level Acoustic Segmentation of Continuous Speech," *Proc.ICASSP 88, IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 429-432.* (Conference in New York, N.Y., April 11-14, 1988.)

Leung, H.C., and V.W. Zue, "Some Phonetic Recognition Experiments Using Artificial Neural Nets," *Proc.ICASSP 88, IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 429-432*, pp. 422-425. (Conference in New York, N.Y., April 11-14, 1988.)

Leung, H.C., and V.W. Zue, "Two-dimensional Characterization of the Speech Signal and Its Potential Applications to Speech Processing," 1st International Conference on Communication Technology, in Nanjing, China, Nov. 1987.

2. Technical Reports:

Lamel, Lori F., "Formalizing Knowledge Used In Spectrogram Reading: Acoustic and Perceptual Evidence from Stops," Technical Report, Research Laboratory of Electronics, Massachusetts Institute of Technology.

3. Presentations:

Invited:

Zue, Victor W., "Phonetically-Based Approach to Automatic Speech Recognition," 21st IBM Computer Science Symposium on "Mechanization of Intelligence and Brain Model," Sendai, Japan, November 1987.

Zue, Victor W., "Discovery of Phonetic Regularities in the Acoustic Signal," CHABA Annual Meeting in Washington, D.C., January 1988

Zue, Victor W., "Speech Recognition Research at MIT," University of Florida, Gainesville, FL, April 1988.

Randolph, Mark. "An augmented Context-Free Parsing Algorithm and its use in Speech Recognition," paper presented at 1988 IEEE workshop on Speech Recognition, New York, May 31- June 3, 1988.

Leung, Hong, "Vowel Recognition Experiments Using Artificial Neural Nets," 1988 IEEE Workshop on Speech Recognition, New York, May 31-June 3.

Contributed:

Lamel, Lori F., "Identification Of Stop Consonants From Continuous Speech In Limited Context: Acoustical Society of America, November 1987.

Leung, Hong. "Recognition of Vowels Using Artificial Neural Networks," Acoustical Society of America, May 1988.

Pitrelli, John. "Factor Analysis for Vowel and Fricative Duration in American English" at Acoustical Society of America, May 19, 1988, Seattle

4. Books (and sections thereof)

Zue, Victor. "Automatic Speech Recognition and Understanding", a chapter in A.I. in the 1980s and Beyond, p. 185-200, Cambridge: MIT Press, 1987.

5. Publications/Patents/Presentations/Honors Report
(Number only)

Papers Submitted to Refereed Journals (and not yet published): 0

Papers Published in Conference Proceedings: 3

Books (and sections thereof) Submitted for Publications: 0

Books (and sections thereof) Published: 1

Patents Filed: 0

Patents Granted: 0

Invited Presentations at Topical or Scientific/Technical Society Conferences: 5

Contributed Presentations at Topical or Scientific/Technical Society Conferences: 3

Honors/Awards/Prizes: 0

Theses: 4

Number of Graduate Students: 6

Number of Post Docs: 0

PHONETIC RECOGNITION USING ARTIFICIAL NEURAL NETWORKS

OBJECTIVES:
Study the basic characteristics of multi-layer perceptrons (MLP). Investigate how its framework can be applied to phonetic recognition when augmented with acoustic-phonetic knowledge.

MOTIVATON:
Many current speech recognition systems are either too rigid for incorporating acoustic-phonetic knowledge or the control strategy is too weak. MLP can potentially bridge the gap between our knowledge and control strategy.

APPROACH:
Use MLP to integrate heterogeneous sources of acoustic and linguistic information.

RESULTS:
- Vowel classification performance comparable to human listeners has been achieved. Accuracy ranges from 66% to 100% under different conditions of the task.

- Performance compares favorably to K-nearest neighbor, a traditional classification technique.

- Speaker adaptation can be achieved by initializing the network properly.

- The network can self-organize its inputs into meaningful phonetic classes.

- Incremental performance improvement on training data provides an effective terminating criterion for training.

- Number of hidden units should depend on the amount of training data.

left context

right context

duration

output from auditory model (synchrony spectrogram)
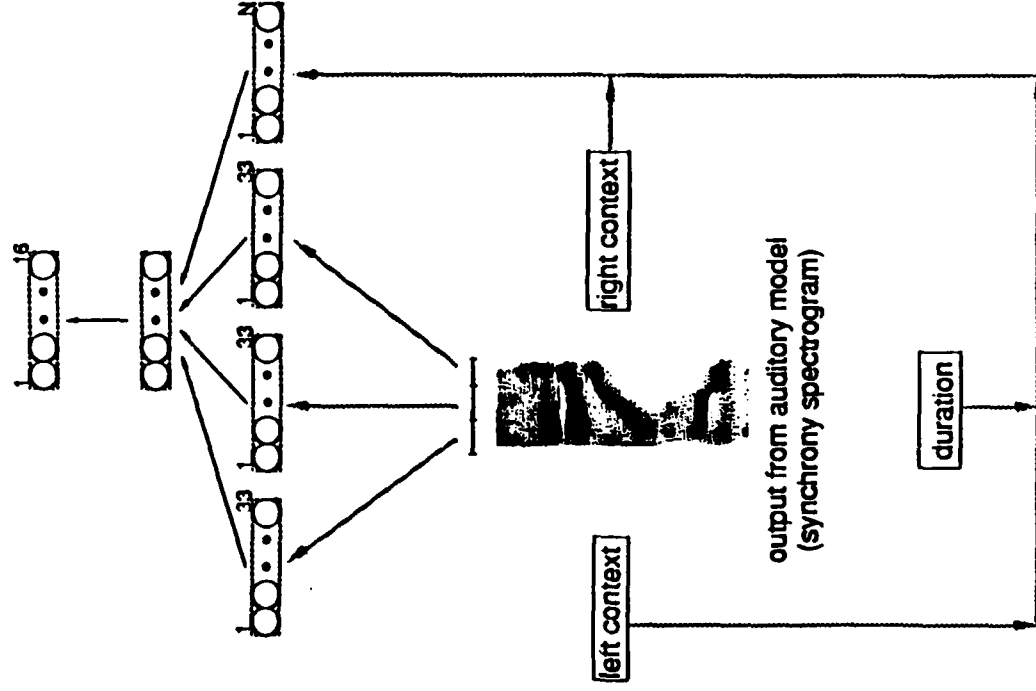
DATABASE: 22,000 vowel tokens, excised from 2,750 continuous sentences spoken by 550 American male and female speakers.

KNOWLEDGE-BASED SPEECH RECOGNITION
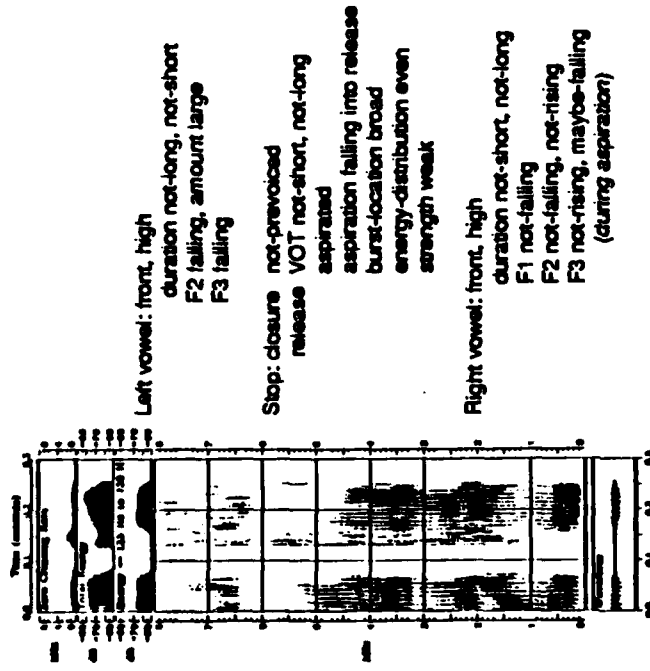
OBJECTIVE: Formalize the knowledge used in speech spectrogram reading by incorporating it in a knowledge-based system

MOTIVATION: Improved understanding of acoustic-phonetics and speech variability is crucial for phonetic recognition

TASK: Identification of stop consonants extracted from continuous speech in a variety of phonemic contexts

ACCOMPLISHMENTS:
- listening experiments indicate that humans can identify stops 85-97% correctly
- spectrogram readers' identification is within 10% of listeners'
- a knowledge-based stop identification system was implemented incorporating acoustic descriptions and reasoning used by spectrogram readers
- system performance is within 10% of spectrogram readers'

APPROACH:
- assess the ability of human listeners to identify stop consonants
- assess the ability of human spectrogram readers to identify stop consonants
- incorporate knowledge obtained from experiments and from spectrogram reading in a knowledge-based system

Left vowel: front, high
   duration not-long, not-short
   F2 falling, amount large
   F3 falling

Stop: closure not-prevoiced
  release VOT not-short, not-long
   aspirated
   aspiration falling into release
   burst-location broad
   energy-distribution even
   strength weak

Right vowel: front, high
   duration not-short, not-long
   F1 not-falling
   F2 not-falling, not-rising
   F3 not-rising, maybe-falling
   (during aspiration)

DISTRIBUTION LIST

	DODAAD Code	
Head Information Sciences Division Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217	N00014	(1)
Administrative Contracting Officer E19-628 Massachusetts Institute of Technology Cambridge, Massachusetts 02139		(1)
Director Naval Research Laboratory Washington, D.C. 20375 Attn: Code 2627	N00173	(1)
Defense Technical Information Center Bldg. 5, Cameron Station Alexandria, Virginia 22314	S47031	(12)